Experimental Emergence of Conventions in Human Dyads:

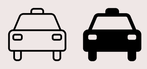
Emergence, stability, and cognitive implications

Oviya Mohan[1] and Dora Biro[1,2]

1. Dept. of Brain and Cognitive Sciences, University of Rochester
2. Dept. of Biology, University of Oxford

omohan@ur.rochester.edu

Introduction: What are Conventions?

- Conventions facilitate solving coordination problems in repeated multi-agent interactions
- Three key properties: e.g. 
  - **Arbitrary** (drive on left vs right)
  - **Efficient** (avoid head-on collisions)
  - **Self-sustaining** (stably maintained in given population)

- The spontaneous emergence of a convention was observed in captive baboons (*Papio papio*)[1] when dyads were tasked with selecting the same color out of two options to receive reward

- What conditions promote the spontaneous formation of conventions?
- How quickly do conventions emerge and how stable are they?
- What aspects of cognition support convention formation and maintenance?

In the Laboratory

Human dyads played a **color-matching game** where different combinations of **seven colors** were presented **pairwise** over 294 trials. Players either received explicit **instructions** (**I**; "choose the same color to score") or **no instructions** (**NI**).

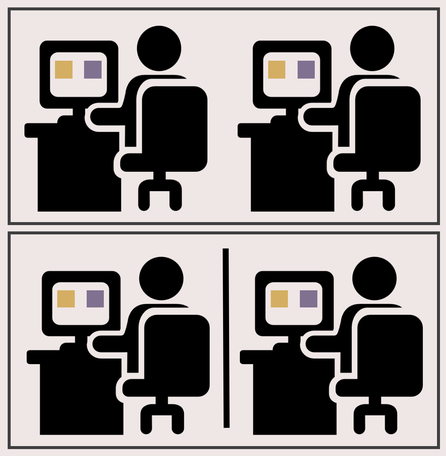
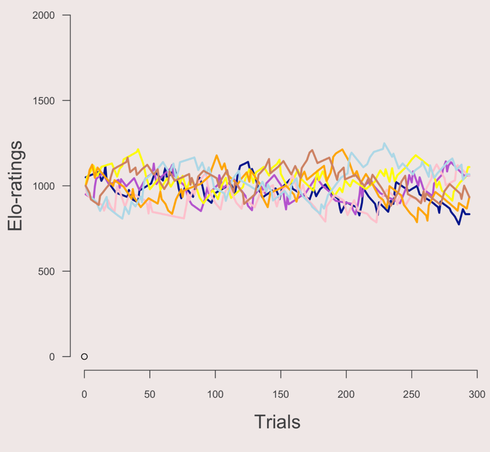


Fig 1. Dyads could either see the partner's screen (top: **transparent condition**, **T**) or had an opaque partition between them (bottom: **opaque condition**, **O**)

Dyads were **not allowed to communicate**!

Conditions	I	NI
O	**I_O**	**NI_O**
T	**I_T**	**NI_T**

Experimentally Induced Conventions



To solve the task, **opaque** dyads had to establish a shared arbitrary ranking (**hierarchy**) of colors.

Transparent dyads could **copy** the partner on each trial OR use a **hierarchy**.

Color rankings were measured continuously via **Elo-ratings**[2].

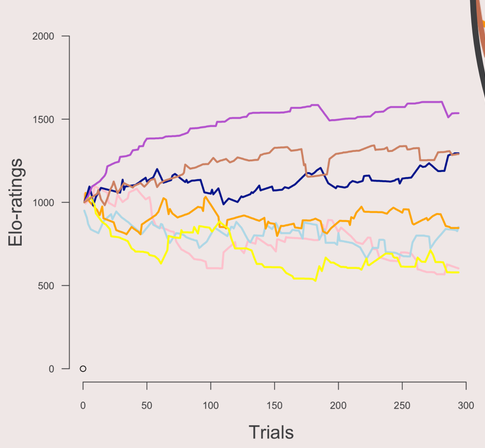
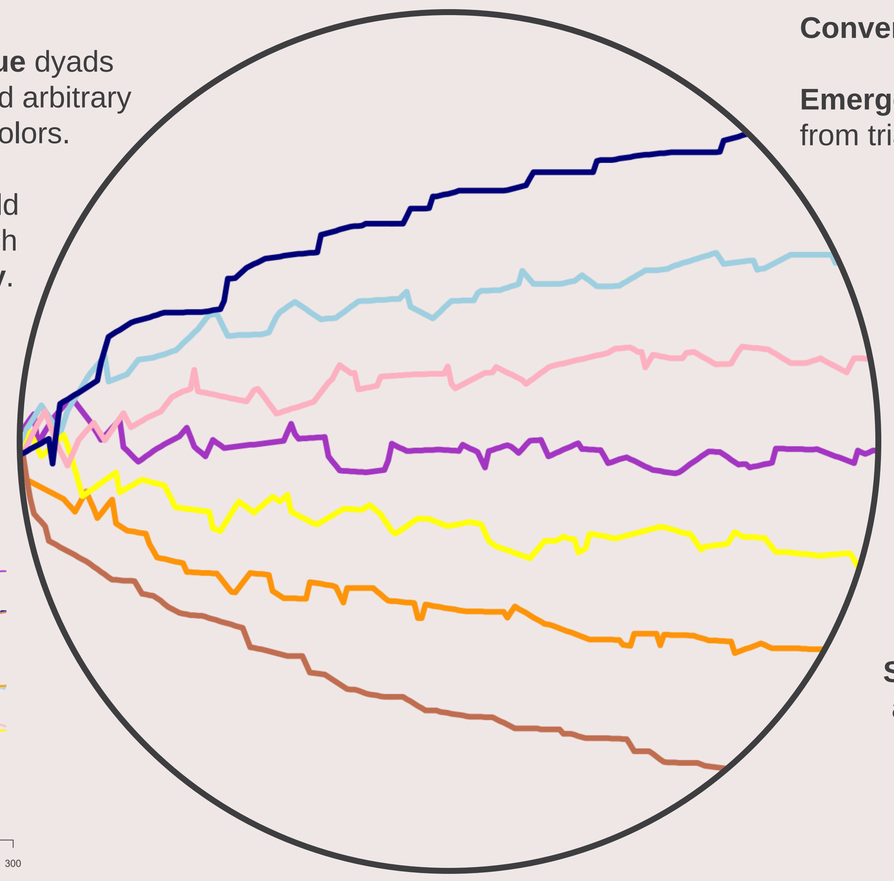


Fig 2. Elo-ratings for individual subjects.
Top: No hierarchy of colors – example from I_T
Right: Partial hierarchy of colors – example from NI_O
Center panel (to the right) : Clear hierarchy of colors – example from I_O



Measuring Conventions

Convention = same hierarchy of colors used by both players

Emergence = convention sustained for criterion number of trials from trial number *n*

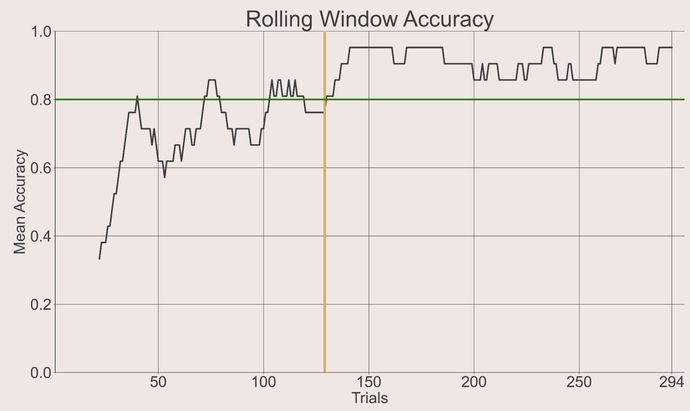


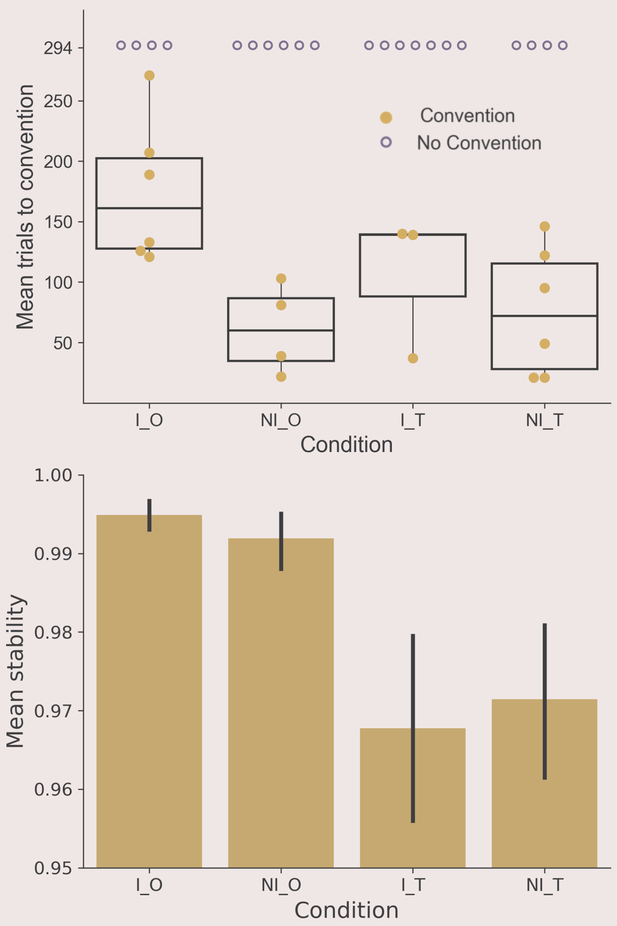
Fig 3. Example of convention emergence (*n* = 133)

Stability = frequency of changes in color rankings once a convention emerges

$$S = \frac{\Sigma_{i=n}^{294}\left(\Sigma_{j=1}^{7}\left|X_{ij} - X_{(i-1)j}\right|\right)}{(294-n)\cdot 7}$$

where X_{ij} is the Elo-ranking of color *j* in trial *i*

Results: Emergence & Stability of Conventions across Conditions



- Conventions emerged across **all conditions**
- **Explicit instructions delayed** the emergence of the convention (Fig. 4 top)
- Conventions were **more stable in the opaque condition** (dyads without visual access) (Fig 4. bottom)

Fig 4. Key results as a function of testing condition

Transparent following Opaque, **O → T**

Opaque following Transparent, **T → O**

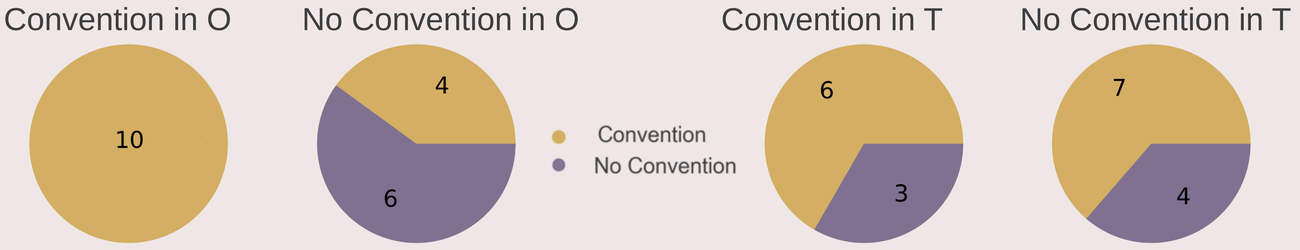


Fig 5. Proportion of dyads using a convention in a second session (immediately following first but with opposite condition for visual access), depending on whether they had a convention in the first session.

Cognitive Implications & Future Directions

- Is a convention the more "efficient" solution? In what ways (e.g. reaction time, memory demands, need for perspective-taking capacities) are conventions more or less efficient?
- What cognitive capacities do subjects employ to establish conventions (e.g. from simple reinforcement tracking to Theory of Mind)?
- Could subjects employ other kinds of conventions (e.g. division of labor as proposer/responder)?
- From dyads to populations: how accurately are conventions transmitted from experienced to naïve subjects?

References

1. Formaux, A., Paleressompoulle, D., Fagot, J., & Claidière, N. (2022). The experimental emergence of convention in a non-human primate. Phil. Trans. R. Soc. B, 377(1843), 20200310. doi: 10.1098/rstb.2020.0310
2. Neumann, C., Duboscq, J., Dubuc, C., Ginting, A., Irwan, A. M., Agil, M., ...Engelhardt, A. (2011). Assessing dominance hierarchies: validation and advantages of progressive evaluation with Elo-rating. Elsevier. Retrieved from http://researchonline.ljmu.ac.uk/id/eprint/3205
3. Image sources: Introduction – CleanPNG, In the laboratory - Flaticon.com